arXiv:0912.2807v2 [physics.soc-ph] 22 Dec 2009

# $p$-th Clustering coefficients $C_p$ and Adjacent Matrix for Networks
# —- Formulation based on String —-

Norihito Toyota
Hokkaido Information University, Ebetsu, Nisinopporo 59-2, Japan
email :toyota@do-johodai.ac.jp

**Abstract**

The phenomenon of six degrees of separation is an old but interesting problem. The considerations of the clustering coefficient reflecting triangular structures and its extension to square one to six degrees of separation have been made[1]. Recently, Aoyama[2] has given some considerations to this problem in networks without loops, using a sort of general formalism, "string formalism". In this article, we describe relations between the string formulation proposed by Aoyama and an adjacent matrix. Thus we provided a reformulation of the string formulation proposed by [2] to analyze networks. According to it, we introduced a series of generalized *q-th* clustering coefficients. The available rules between diagrams of graphs and formulae are also given based on the formulation. Next we apply the formulation to some subjects in order to mainly check consistency with former studies. By evaluating the clustering coefficient for typical networks studied well earlier, we confirm a validity of our formulation. Lastly we applied it to the subject of two degrees of separation.

**keywords:** Six Degrees of Separation, String, Clustering Coefficient, Degree Distribution, Generalized Clustering Coefficient

## 1 Introduction

In 1967, Milgram has made a great impact on the world by advocating the concept "six degrees of separation" by a celebrated paper [3] written based on an social experiment. Six degrees of separation indicates that people have a narrow circle of acquaintances. A series of social experiments made by him and his joint researcher[4],[5] suggested that all people in USA are connected through about 6 intermediate acquaintances. Their studies were strongly inspired by Pool and Kochen's study [6]. At the time, however, numerically detailed studies could not be made because computers and important concepts, such as the clustering coefficient needed for a network analysis nowadays, have not yet developed sufficiently.

One of the most refined models of six degrees of separation was formulated in work of Watts and Strogatz[7],[8]. Their framework provided compelling evidence that the small-world phenomenon is pervasive in a range of networks arising in nature and technology, and a fundamental ingredient in the evolution of the World Wide Web. Another is the scale free networks proposed by Barabasi et al.[9], [10]. Many empirical networks are characteristic future of scale free [11],[13],[14],[15]. In spite of furthermore exploring of six degrees of separation[16],[17], they do not examine closely Milgram's original findings by their model, especially how influence can the clustering coefficient proposed in the paper [7] has. We have made some study of it in our previous paper [18] by imposing a homogeneous hypothesis on networks. As a result, we found that the clustering coefficient has not any decisive effects on the propagation of information on a network and then information easily spread to a lot of people even in networks with relatively large clustering coefficient under the hypothesis; a person only needs dozens of friends. Moreover we devoted deep study to the six degrees of separation based on some models proposed by Pool and Kochen [6] by using a computer, numerically[19]. In the article, we estimated the clustering coefficient along the method developed by us [18] and improved our analysis of the subject through marrying Pool and Kochen's models to our method introduced in [18]. As a result, it seems to be difficult that six degrees of separation is realized in the models proposed by Pool and Kochen[6] on the whole.

If the network of human relations has a tree structure without loops, a person connects new persons in power of average degree, when he(she) follows his(her) acquaintances step by step on his(her) network of human relations. Then the phenomenon of six degrees of separation is not so amazing, if a person has a few hundred acquaintances. A question is that networks of general human relations include some loop structures. This structures decrease the number of new persons that connected with him(her) when he(she) follow his(her) acquaintances step by step. One of indices characterizing loop structures is the clustering coefficient. Thus it will be important to investigate the effect of the clustering coefficient and degree distribution on the six degrees of separation. It is, however, difficult to investigate the influence of general loop structures. There are only a little research focused on the effect of loop structures.

Newman first studied the influence of loop structures in a network on the subject[1]. The study is so stimulating but only triangle structures and quadrilateral structures on networks were considered. It seems to be difficult to generalize his framework to $q$-polygon. Recently Aoyama proposed the string formulation on the subject[2]. The idea inspired our study in this article, greatly. Unfortunately he considered only tree approximation in the structure of networks. Since he deals with mainly scale free networks, the approximation is valid up to a point.

In this article we pursue the string formulation and try to discuss the influence of general loops to six degrees of separation. One of the aims in this article is to reformulate the string formulation based on an adjacent matrix. We can systematically analyze general networks with arbitrary loop structures by the reformulation. Next is to check the results derived from it are consistent with results studied so far. Then we apply it to the problem of two degrees of separation as a first step so that future prospect for the problem are opend.

The plan of this article is as follows. After introduction, we reformulate the string formulation by an adjacent matrix in the following section 2. First we explain some notations used in this article and reformulate the string formulation by using an adjacent matrix. Here we introduce $R$ matrix that play central role in our formalism. According the formalism, we introduce generalized $q$-th clustering coefficients as well as the usual global one. Next we give a diagrammatic interpretation for every term appearing in the expansion of the power of $R$ matrix like Feynman diagram [21] in quantum field theory. In the next section 3, we evaluate some clustering coefficients on some typical networks and discuss the consistency with results investigated by now. We shall discuss the phenomenon of six degrees of separation in the section 4. Since any reliable conditions have not been given for $p$-th degrees of separation in networks with strongly connected components yet, we can not provide full discussions. Thus we discuss two degrees of separation and compare the results to those given by Aoyama[2] in scale free networks. We find that the result is a little different from Aoyama's one[2]. The analysis of an adjacent matrix is seemed to support rather our results. We shall devote discussion on six degree of separation based on the formulation in a subsequent article[20]. The last section 5 is devoted to summary.

# 2 String Formulation and Adjacent Matrix

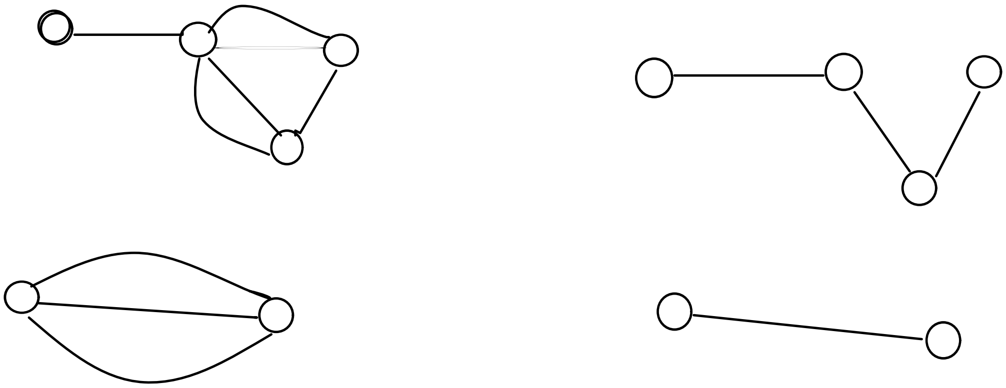
We, basically, follow the string formulation introduced by Aoyama [2].

## 2.1 Notations

In this section we describe notations used in this article. We consider a string-like part of a graph with connected $j$ vertices and call it "j-string". $N$ is the number of vertices in a considering network and $S_j$ is the number of j-string in the network. (Note that $S_j$ in this article is $N$ times larger than $S_{j-1}^{Aoyama}$ defined by Aoyama[2].) By definition, $S_1 = N$ and $S_2$ is the number of edges in the network. $\bar{S}_j$ is the number of nondegenerate j-string where a nondegenerate string is defined as strings without any multiple edges and/or any loops in the subgraphs as seen in Fig.1. We, however, define that the nondegenerate string contains strings homeomorphic to a circle.

We call strings without any loops open string and strings homeomorphic to a circle closed string. Thus we consider closed strings and open strings in this article.

It is so difficult to calculate $S_j$ and $\bar{S}_j$, generally. Aoyama has calculated $S_j$ up to $j = 7$ but did not supply the explicit in his article[2]. His article says that it would needs dozens of papers if the full expression is described. It would be maybe impossible to calculate $S_j$ and $\bar{S}_j$ with $j > 7$ explicitly at the present moment.



(a)Degenerate strings (b)Nondegenerate strings

Figure 1: Two types of strings

## 2.2 Clustering Coefficient

By using the string formulation, we can defined the usual clustering coefficient which essentially counts the number of triangular structures in a network. Let $\Delta_q$ be the number of polygons with $q$ edges in a network. Following Newman [1], the usual global clustering coefficient $C_{(3)}$ is given by

$$C_{(3)} = \frac{6\times\ number\ of\ triangles}{number\ of\ connected\ triplets} = \frac{6\Delta_3}{\bar{S}_3}. \tag{1}$$

We introduce some indices uncovering properties of general polygon structures except for triangle structures in a network. From the expression of Eq.(1), we can generalize it to *q-th* clustering coefficient $C_{(q)}$ straightforwardly;

$$C_{(q)} = \frac{2p\times\ number\ of\ polygons}{number\ of\ connected\ p\text{-}plets} = \frac{2p\Delta_p}{\bar{S}_p}. \tag{2}$$

## 2.3 Adjacent Matrix Formulation

We reformulate $C_{(q)}$ introduced in Eq.(2) by utilizing an adjacent matrix $A = (a_{ij})$. Generally the powers, $A^2$, $A^3$, $A^4$, $\cdots$ of adjacent matrix $A$ give information as to respecting that a vertex connects other vertices through $2, 3, 4, \cdots$ intermediation edges, respectively. The information of the connectivity between two vertices in $A^n$ contains multiplicity of edges, generally. For resolving the degeneracy, we introduce a new se ries of matrices $R^n$ which give information as to respecting that a vertex connects other vertices through $n$ intermediation edges without multiplicity. We can find it by the following formula[22];

$$R^n = \sum_{i_1,\cdots,i_{n-1}} a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i-1,i_n} \frac{\prod\limits_{i_k,i_j,i_k-i_j>1}^{n} (1-\delta_{i_k i_j})}{(1-\delta_{i_0 i_n})}. \tag{3}$$

where the product of $(1-\delta_{i_k i_j})$ of the numerator has the role of protecting of degeneracies from strings and $(1-\delta_{i_0 i_n})$ of the denominator is, however, needed to keep a closed string.

This expression has $(n-1)$ply loops in a computer program and so it is almost impossible to calculate $R^n$ within real time for large $N$ when the rank of $R$ is $N$. The expansion of Eq.(3) has $2^{n(n-1)/2}$ terms formally. This value is 32768 for $n = 6$ that is needed in the analysis of six degrees of separation as will be discussed in the later section. Though many terms really vanish, $R^6$ has still so complex expression. We only give the expressions of $R^1 \sim R^5$ here and will give that of $R^6$

in Appendix;

$$
\begin{aligned}
[R^2]_{if} &= [A^2]_{if} - [A^2]_{ii}\delta_{if} = [A^2]_{if} - G_{if}, \\
[R^3]_{if} &= [A^3]_{if} - \{G, A\}_{if} + a_{if}, \\
[R^4]_{if} &= [A^4]_{if} - \{G, A^2\}_{if} + \big\{A, diag(A^3)\big\}_{if} + 2[A^2]_{if} + [G^2 - G - AGA]_{if} + 3a_{if}[A^2]_{if} \\
[R^5]_{if} &= [A^5]_{if} - \big\{A, diag(A^4)\big\}_{if} - \{G, A^3\}_{if} - \big\{A^2, diag(A^3)\big\}_{if} + 3\big([A^2]_{if}\big)^2[A]_{if} \\
&\quad + 3[A^3]_{if}[A]_{if} + 2\{G^2, A\}_{if} + [GAG]_{if} - 6\{G, A\}_{if} - \{AGA, A\}_{if} + 3[A^3]_{if} \\
&\quad + \big\{A, diag(AGA)\big\}_{if} + 2[diag(A^3G)]_{if} - [A \cdot diag(A^3) \cdot A]_{if} - [diag(A^3)]_{if} \\
&\quad + 3\sum_k a_{ik}a_{kf}\Big([A^2]_{kf} + [A^2]_{ik} - \delta_{if}[A^2]_{kf}\Big) + 4a_{if}\big(1 - a_{if}\big),
\end{aligned} \tag{4}
$$

where suffix is abbreviate in trivial cases and $\{\cdot,\cdot\}$ means the anticommutation relation; $\{A, B\} = AB - BA$. $diagA$ indicates the diagonal matrix whose elements are the diagonal elements of $A$, and $G$ is the diagonal matrix defined by

$$
G \quad = \quad \begin{bmatrix} k_1 & 0 & 0 & \cdots \\ 0 & k_2 & 0 & \cdots \\ 0 & 0 & k_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{5}
$$

where $k_i$ is the degree of vertex $i$.

By using $R^n$, $\bar{S}_p$ and generalized $p$-th clustering coefficient $C_{(p)}$ are given by

$$
\bar{S}_p = \sum_{i,j}(R^{p-1})_{ij}/2, \tag{6}
$$

$$
C_{(p)} = \frac{\mathrm{Tr}R^p}{\sum_{i,j} R^{p-1}}, \tag{7}
$$

where denominator and numerator indicates the contribution from open strings and a closed string, respectively. Thus usual clustering coefficient $C_{(3)}$ becomes

$$
C_{(3)} = \frac{\mathrm{Tr}R^3}{\sum_{i,j}(A^2)_{ij} - (A^2)_{ij}\delta_{ij}} = \frac{\mathrm{Tr}A^3}{||A|| - \mathrm{Tr}A^2}. \tag{8}
$$

where we introduced a new symbol $||\cdots||$ which denotes $||A|| \equiv \sum_{i,j} A_{ij}$.

## 2.4 Diagrammatic Interpretation

The expression of $R^n$ is rapidly complicated as $n$ increases. We, however, notice that every term appearing in the expansion of $R^n$ closely corresponds to a certain graph[22]. So if a certain graph is given, we can write out the expression corresponding to it like Feynman's rule in quantum field theory[21]. We describe the diagrammatic construction of $R^n$ based on the relation between both.

1. Draw all graphs with $n$ edges including degenerate graphs except for closed string.

2. Assign sign $(-1)^{n-1+v}$ for every graph where $v$ is the number of vertices included in the graph.

3. Calculate degeneracy index $m$ defined in the following Eq.(9) and it is the coefficient of the term corresponding to the graph;

$$
m = \prod_{i,k_i \neq 1} [\frac{k_i - 1}{2}]_{Ga}, \tag{9}
$$

where $[\ \ ]_{Ga}$ means Gauss symbol. The coefficient supply essentially the number of Euler paths in a graph. We do not distinguish a path and its opposite path and start from vertex with hight degree in the cases that there is not any odd vertex in the graph.

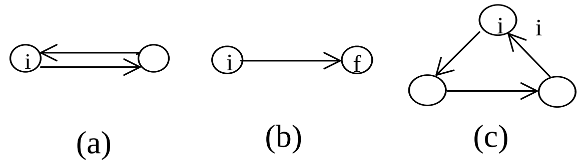



Figure 2: Typical diagrams@I

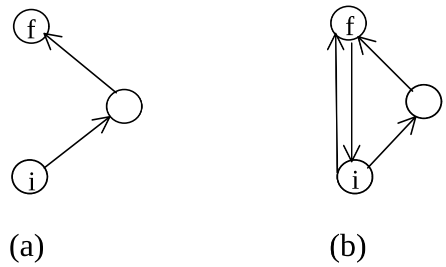



Figure 3: Typical diagrams II

Thus the subject of finding the expression of $R^n$ reduces to the estimation of diversity of graph with certain constant number of edges and the number of Euler paths in the graph.

We describe the typical relations between expressions and graphs. The typical graphs are given in Fig.2 and Fig.3 where the arrows indicate the order in Euler paths and i and f shows an initial vertex and a final one, respectively. These graphs correspond to the following expressions;

$$Figure2(a) = k_i, \tag{10}$$
$$Figure2(b) = a_{if}, \tag{11}$$
$$Figure2(c) = (A^3)_{ii}, \tag{12}$$
$$Figure3(a) = (A^2)_{if}, \tag{13}$$
$$Figure3(b) = a_{ij}(A^2)_{if}. \tag{14}$$

It is assumed that the right vertex has not any other edges in (a) of Fig 2. The generalization of Eq.(12) to polygons with $n$ vertices is easily given by $(A^n)_{ii}$. Eq.(13) is also straightly extended to $(A^n)_{if}$ in the cases that $n$ vertices are linearly connected.

Degenerate multigraphs also reduce multigraphs with little multiplicity as shown in Fig.4. The left graph comes down to the corresponding expression $a_{if}$ in (a) of Fig.4. Since these are only useful relations between graphs and expressions, these graphs should be independently considered in the step 1 in the diagrammatic construction. These correspondence may be rather effective in estimating explicit expression of $R^6$.

## 2.5 An Example

We give an example of the diagrammatic construction of $R^n$. As a nontrivial case, we consider the construction of $R^4$ which is constructed from graphs with four edges. All topologically independent graphs with four edges are given in Fig.5. For every graph, signs and coefficients derived from step

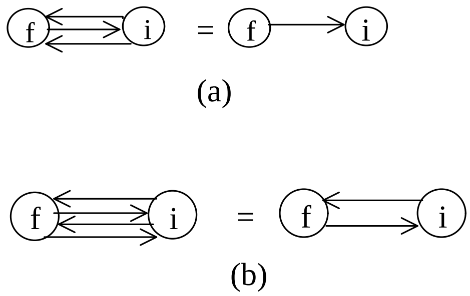



Figure 4: Reducing of diagrams

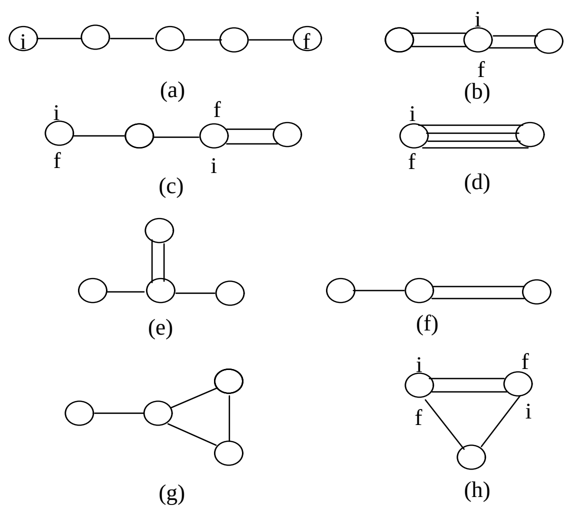



Figure 5: Topologically independent graphs with four edges

2 and 3 in the diagrammatic construction are as follows;

$$\begin{aligned}
Figure5(a) &= [A^4]_{if}, && (15)\\
Figure5(b) &= [G^2]_{if}, && (16)\\
Figure5(c) &= -\{A^2, G\}_{ii}, && (17)\\
Figure5(d) &= -[G]_{if}, && (18)\\
Figure5(e) &= -[AGA]_{ij}, && (19)\\
Figure5(f) &= 2[A^2]_{if}, && (20)\\
Figure5(g) &= A^3_{ii}a_{if} + A^3_{ff}a_{if}, && (21)\\
Figure5(h) &= 3a_{ij}[A^2]_{if}. && (22)
\end{aligned}$$

It is easy to confirm that the sum from Eq.(15) to Eq.(22) give $R^4$ in Eq.(4).

# 3 Clustering Coefficient of Some Networks

In this section we calculate $C_{(3)}$ in some typical networks by using the formulation given in the previous section. Thus we investigate consistency of our formulation with results observed so far. We adopt the configuration model [23].[24],[25] in producing networks. The model can systematically produce networks with arbitrary degree distribution. But the networks produced by the model are degenerate multigraphs, generally. We modify it a little to produce networks without multiplicity. Since it is not essential in this article, we omit the technical details of it.

First we study random networks[26] where the degree distribution is Poisson distribution[9]. The $N$ dependence on $C_{(3)}$ given by computer simulations is shown in Fig.6. Theoretically it

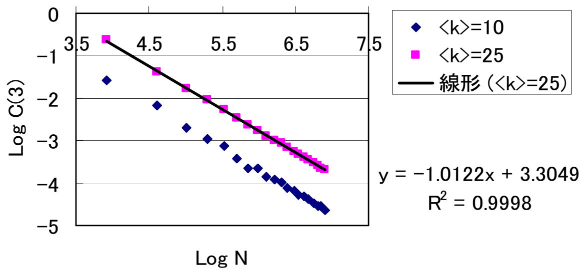



Figure 6: The scaling of $C_{(3)}$ in random graphs with Poisson distribution

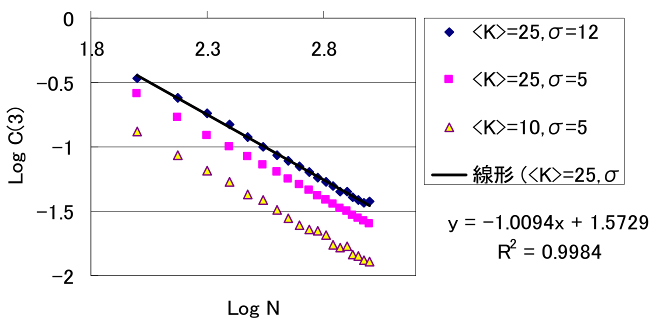



Figure 7: The scaling of $C_{(3)}$ in graphs with Normal distribution

is shown that the clustering coefficient behaves as $C_{(3)} = \langle k \rangle / N$ in random networks where $\langle k \rangle$ indicates average degree [11]. The log-log plot in Fig.6 shows linear behavior with slope nearly 1. Thus $N$ dependence of $C_{(3)}$ is consistent with observations so far.

Moreover we can observe similar scaling law in networks with the normal distribution in degree distribution. As Fig.7, all the slopes of the log-log plot indicating $N$-$C_{(3)}$ relation are also nearly 1 for various values of $\langle k \rangle$ and the standard deviation $\sigma$. The delicate difference of the values of the slopes was discussed in [22].

It is known that the clustering coefficient follows the scaling law $C_{(3)} \sim N^{-0.75}$ for scale free networks[11] with $\langle k \rangle = 4$. Our simulation results for scale free networks that have the average degree 4 are given by Fig. 8. The slope, about -0.73, of the approximative line in Fig.8 shows that the above result almost agrees with current one. This also strengthens the validity of presented formulation.

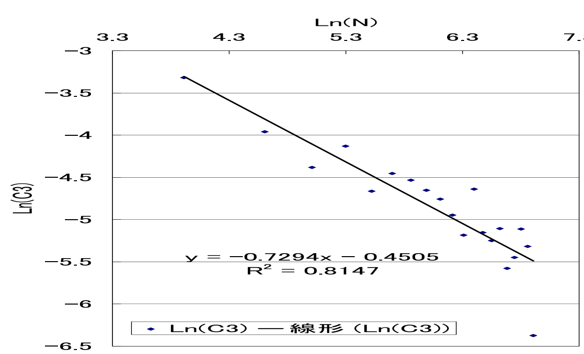



Figure 8: The scaling of $C_{(3)}$ in scale free networks with $\langle k \rangle = 4$

# 4 Application to Two Degrees of Separation

Aoyama has proposed a condition, so-called Milgram Condition, for *p-th* degrees of separation[2];

$$M_p \equiv \frac{\bar{S}_p}{N} \sim O(N). \tag{23}$$

For six degrees of separation, we obtain from Eq.(6)

$$\bar{S}_7 = \sum_{i,j} (R^6)_{ij}/2. \tag{24}$$

Before evaluating $R^6$ for six degrees of separation, we as a first step study two degrees of separation to check consistency in this article. Since 2-string cannot have any loops, we have only to consider graphs by the tree approximation in the study of two degrees of separation.

Fig. 9 and Fig.10 show $M_2$ and $N$ in Eq.(23) for random networks with Poisson distribution and networks with the normal distribution in degree distribution. From the figures, we observe a relation $M_2 \sim const.$, independently of $N$. The linear lines show $N$ in the right hand side of Eq.(23). Though one intersection point in the Fig.9 and Fig.10 only makes Milgram Condition, $M_2 > N$ means that two degrees of separation is realized. It is getting quite difficult that the Milgram Condition is satisfied for small $\langle k \rangle$. Though $N \leq 1000$ in our simulations, we can also speculate where is the intersection point between two lines for large $\langle k \rangle$ due to $M_2 \sim const.$ for Normal distribution. Moreover we notice the constant values are almost proportional to $\langle k \rangle^2$. It is natural that such situation occurs because loop structures can be neglected in two degrees of separation. Thus we can approximately estimate the critical point that the Milgram Condition is satisfied for arbitrary $\langle k \rangle$. In this connection, we obtain $\langle k \rangle \sim$ (a few $\times 10^4$) for $N \sim$ (a few $\times 10^8$) which is about the population of USA, in a random network. We think that the value of $\langle k \rangle$ is as large as it is unrealistic, because some social researchers estimated the average number of acquaintances of a person is 290 [27],[28],[29].

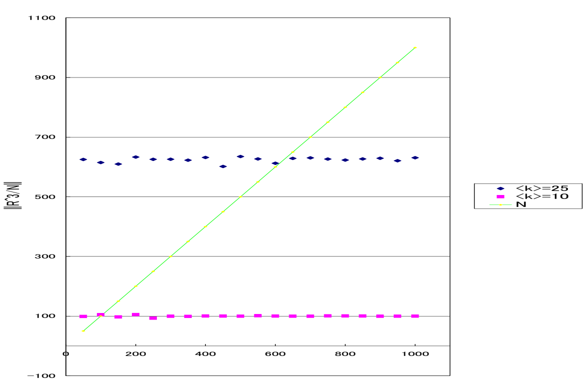



Figure 9: $M_2$ with $\langle k \rangle = 10$ and $\langle k \rangle = 20$ for Poisson distribution

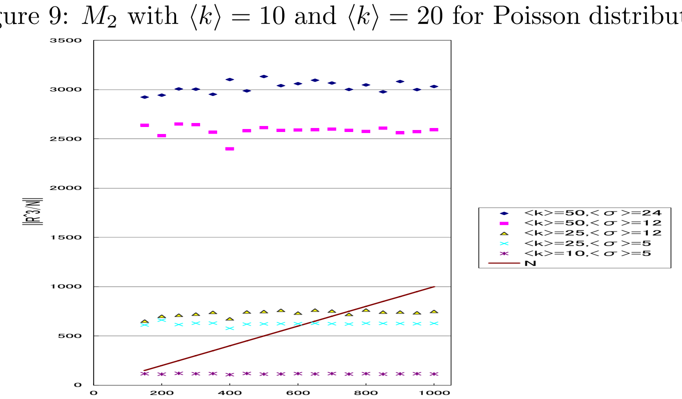



Figure 10: $M_2$ with $(\langle k \rangle, \sigma) = (10.5), (25.5), (25, 12), (50.12), (50, 24)$ for Normal distribution

Figure 11: $M_2$ and $N$ for $\gamma = 2, 3, 4$ in SF networks

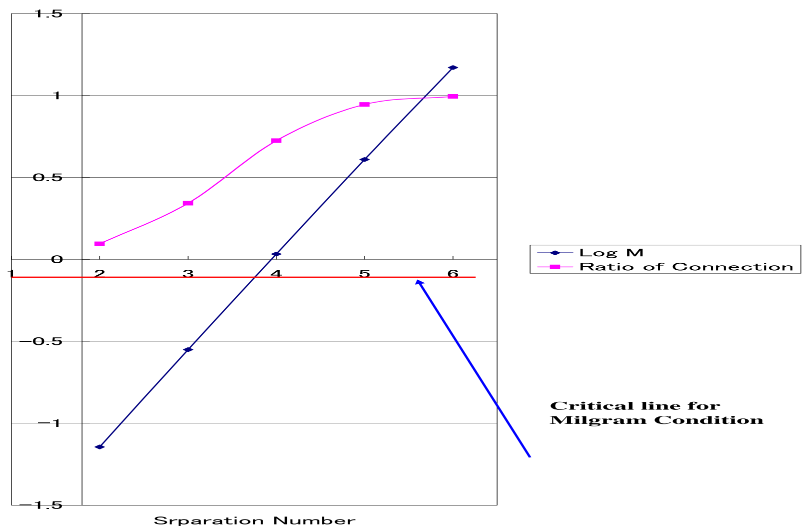



Figure 12: $M_2/N$ and Connection Ratio v.s. degree number of separation for SF network with $\gamma = 2$

The behavior of $M_2$ is rather different from the Normal and Poisson random networks for scale free networks with degree distribution $P(k) \sim k^{-\gamma}$. $\langle k \rangle$ depends on the index $\gamma$ in the scale free networks produced by the configuration model. As $N$ increases, so $M_2$ increases quickly than $N$ as is shown in Fig.11. The increasing rate is larger for the smaller index . Aoyama[2] has pointed out that critical index is $\gamma = 2$ in two degrees of separation. Fig..11, however, shows that $\gamma \leq 2$ can not realize the two degrees of separation. The result is different from Aoyama's one.

For $\gamma = 2$, $M_p/N$ and the ratio of the zero components in $A^2$ in our formalism are given in Fig.12 where $p$ means the $p-$th degrees of separation ("separation number"). Evaluating the ratio of the zero components, $\gamma \leq 2$ can not realize two degrees of separation but rather four degrees of separation is realized where 75 percent of all the nodes are connected. From (23), the critical velue of $\log_{10} M_2/N$ would be an upper areas of the line a little smaller than zero. The $M_p/N$ satisfying the Milgram condition in Fig.12 also support this results. So, our formalism is thought to be available in order to analyze degree numbers of separation. We shall devote discussion on six degree of separation based on the formulation in a subsequent article[20].

## 5 Summary

In this article, we provided a reformulation of the string formulation proposed by [2] to analyze networks. Fusing adjacent matrix into the formalism, we reformulate the string formalism. According to it, we introduced a series of generalized *q-th* clustering coefficients. Their function is

not yet considered in this article and left as a theme for research in the future. Then we introduce the $R$ matrix in the formalism developed in this article instead of $A$. The power of $R$ plays central role in the analysis of this article. Every term of the expansion of $R^n$ can be also interpreted graphically and it would make a projection for the future in the estimation of $R^n$ that has complex expressions for large $n$.

On the latter half of this article, we apply the formulation to some subjects in order to mainly check consistency with former studies. We first evaluated the clustering coefficient for typical networks studied well earlier. We could confirm a validity of our formulation by these in some degree. Lastly we applied the formulation to the subject of two degrees of separation. We find that the result is a little different from Aoyama's one[2]; the separation number is not two but four at $\gamma = 2$. It is noticed that by analyzing the number of zero-components in $A^2$, our results are rather supported.

The following problems are yet left in future:

1. Finding explicit expressions of $R^n$ for arbitrary $n$ by applying our formalism, especially diagrammatic construction. Then finding a general formula for arbitrary $n$ from the series of expression.

2. Revealing relations between *p-th* degrees of separation and $N$, $\langle k \rangle$ or $< k^n >$. More definitely, discovering the relations between $p$ and $N$, $\langle k \rangle$ or $< k^n >$.

3. Revealing relations between *p-th* degrees separation and various loop structures, especially $C_{(q)}$.

The last one will be discussed in the subsequent paper [20]. This article give a sort of general formalism to investigate above problems, including preliminary studies of them.

# References


[1] M.E.J.Newman,"Ego-centered networks and the ripple effect or why all your friends are wired", Social Networks 25 (2003) p.83;arXiv. cond-mat/0111070

[2] H. Aoyama, "Six degrees of separation; some caluculation", SGC library65, " Introduction to Network Science", (2008) in Japanese; H, Aoyama, Y.Fujiwara, H, Ietomi, Y. Ikeda and W.Soma "EconoPhysics",Kyouritu Shuppan 2008

[3] S. Milgram, "The small world problem", Psychology Today 2, 60-67 (1967)

[4] J. Travers and S. Milgram, "An Experimental Study of the Small World Problem", Sociometry 32, 425 (1969)

[5] C. Korte and S. Milgram, "Acquaintance edges between White and Negro populations: Application of

[6] I.S. Pool and M. Kochen, "Contacts and Influence", Social Networks, 1(1978/1979)5-51(This paper was actually written in 1958)

[7] D. J. Watts@and S. H. Strogatz, "Collective dynamics of 'small-world' networks",@Nature,393, 440-442(1998)

[8] D. J. Watts, "Six degree– The science of a connected age", W.W. Norton and Company, New York (2003)

[9] A.-L.Barabasi and R.Albert, "Emergence of scaling in random networks", Science, 286, 509-512(1999)

[10] A.-L.Barabasi and R.Albert, "edgeed: The New Science of Networks", Perseus Books Group (2002) edgeed: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life Plume ; ISBN: 0452284392 ; Reissue (2003/04/29)

[11] R.Albert and A-.L. Barabasi, "Statistical Mechanics of complex networks",Rev. Mod. Phys. 74, 47-97(2002)

[12] J.S. Kleinfield, "The small world problem", Society 39(2) pp.61-66(2002): "COULD IT BE A BIG WORLD? ", http://www.uaf.edu/northern/big_world.html

[13] M.E.J. Newman, A.-L.Barabasi and D. J. Watts, "The Structure and Dynamics of Networks", Princeton Univ. Press, 2006 @

[14] S. N. Dorogovtsev, A.V. Goltsev and J.F.F. Mendes, "Pseudo fractal scale-free web", Phys. Rev. E.65, 066122(2002)

[15] S. N. Dorogovtsev and J.F.F. Mendes, "Evolution of Networka", Oxford Univ. Press, Oxford(2003)

[16] D. J. Watts et al., Small World Project-Columbia University. http://small world.columbia.edu/

[17] P.S.Dodds, R.Muhamad and D.J. Watts, "An Experimental Study of Research in Global Social Networks", Science 301, pp.827-829: http://small world.columbia.edu/images/dodds2003pa.pdf (2003)

[18] N. Toyota, "Some Considerations on Six Degrees of Separation from A Theoretical Point of Viewh, arXiv:0803.2399

[19] N. Toyota, "Comments on Six Degrees of Separation based on the le Pool and Kochen Modelsgendary", arXiv:0905.4804

[20] N. Toyota and T. Sakamoto, to be appeared.

[21] M. E. Peskin and V. Schroeder, "An introduction to Quantum Field Theory", Westview (1995)

[22] N. Toyota, IEICE Thecnical Report, "String Formalism for $p$-Clustering Coefficient-Toward Six Degrees of Separations",NLP2009-49(2009)

[23] A.Bebessy, P.Bebessy and J. Komlos, Stud,. Sci., Math. Hangary, 7343- 7353 (1972)

[24] E.A.Bender and E.R. Candield, J. Comb. Theory A. 24. 296-307 (1978)

[25] M. Molloy and B. Reed, Comb., Prob. and Compt. 6. 161-179 (1995); 7. 295-305 (1998

[26] P. Erdos and A. Renyi," On random graphs I", Publicationes Mathematicae Debrecen6, 290-297, 1959

[27] P.D.Killwoth,E.C.Johnsen, H.R.Bernard, G.A.Shelley and "Estimating the size of personal networks", Social Networks 12,289-312 (1990)

[28] H.R.Bernard, E.C.Johnsen, P.D.Killwoth and S. Robinson, " Estimating the size of average personal network and of an event population; Some empirical results", Social Science Research 20, 109-1211991)

[29] H.R.Bernard, P.D.Killwoth, E.C.Johnsen, and C.McCarty, " Estimating the ripple effect of a disaster", Connections 24(2), pp.16-22(2001)

## A $R^6$ and Tr$R^n$

In this appendix we give an explicit expression $R^6$ and the expressions of Tr of $R^n$ for $n = 1 \sim 6$. $R^6$ is obtained after straightforward but long tedious calculation. We divide it into the following four parts to brighten the prospects of the caluculation.

$$
\begin{aligned}
[R^6]_{if} =& \sum_{j,k,l,m,n} a_{ij}a_{jk}a_{kl}a_{lm}a_{mn}a_{nf} \times \Delta_{ik}\Delta_{jl}\Delta_{km}\Delta_{ln}\Delta_{mf}\Delta_{il}\Delta_{jm}\Delta_{kn}\Delta_{lf}\Delta_{im}\Delta_{jn}\Delta_{kf}\Delta_{in}\Delta_{jf} \\
=& \sum_{j,k,l,m,n} a_{ij}a_{jk}a_{kl}a_{lm}a_{mn}a_{nf} \times \Delta_{ik}\Delta_{jl}\Delta_{km}\Delta_{ln}\Delta_{mf}\Delta_{il}\Delta_{jm}\Delta_{kn}\Delta_{lf}\Delta_{im}\Delta_{jn}\Delta_{kf} \\
& - \sum_{k,l,m,n} a_{if}a_{fk}a_{kl}a_{lm}a_{mn}a_{nf} \times \Delta_{ik}\Delta_{fl}\Delta_{km}\Delta_{ln}\Delta_{mf}\Delta_{il}\Delta_{kn}\Delta_{im} \\
& - \sum_{j,k,l,m} a_{ij}a_{jk}a_{kl}a_{lm}a_{mi}a_{if} \times \Delta_{ik}\Delta_{jl}\Delta_{km}\Delta_{li}\Delta_{mf}\Delta_{jm}\Delta_{lf}\Delta_{kf} \\
& + \sum_{k,l,m} a_{if}a_{fk}a_{kl}a_{lm}a_{mi}\Delta_{ik}\Delta_{fl}\Delta_{km}\Delta_{li}\Delta_{mf}, \\
\equiv& R^6[1]_{if} + R^6[2]_{if} + R^6[3]_{if} + R^6[4]_{if},
\end{aligned} \tag{25}
$$

where $\Delta_{ik} = 1 - \delta_{ik}$. Furthemore we divide $R^6[1]_{if}$ into the following four parts to brighten the prospects of the caluculation.

$$
\begin{aligned}
R^6[1]_{if} =& \sum_{j,k,l,m,n} a_{ij}a_{jk}a_{kl}a_{lm}a_{mn}a_{nf} \times \Delta_{ik}\Delta_{jl}\Delta_{km}\Delta_{ln}\Delta_{mf}\Delta_{il}\Delta_{jm}\Delta_{kn}\Delta_{lf}\Delta_{im}\Delta_{jn}\Delta_{kf} \\
=& \sum_{j,k,l,m,n} a_{ij}a_{jk}a_{kl}a_{lm}a_{mn}a_{nf} \times \Delta_{ik}\Delta_{jl}\Delta_{km}\Delta_{ln}\Delta_{mf}\Delta_{il}\Delta_{jm}\Delta_{kn}\Delta_{lf}\Delta_{jn} \\
& - \sum_{j,k,l,n} a_{ij}a_{jk}a_{kl}a_{li}a_{in}a_{nf} \times \Delta_{ik}\Delta_{jl}\Delta_{ln}\Delta_{if}\Delta_{kn}\Delta_{lf}\Delta_{jn}\Delta_{kf} \\
& - \sum_{j,l,m,n} a_{ij}a_{jf}a_{fl}a_{lm}a_{mn}a_{nf} \times \Delta_{if}\Delta_{jl}\Delta_{fm}\Delta_{ln}\Delta_{il}\Delta_{jm}\Delta_{jn} \\
& + \sum_{j,l,n} a_{ij}a_{jf}a_{fl}a_{li}a_{in}a_{nf}\Delta_{if}\Delta_{jl}\Delta_{ln}\Delta_{jn}\Delta_{mf}, \\
\equiv& R^6[1,1]_{if} + R^6[1,2]_{if} + R^6[1,3]_{if} + R^6[1,4]_{if}.
\end{aligned} \tag{26}
$$

The four terms are respectively expressed as follows;

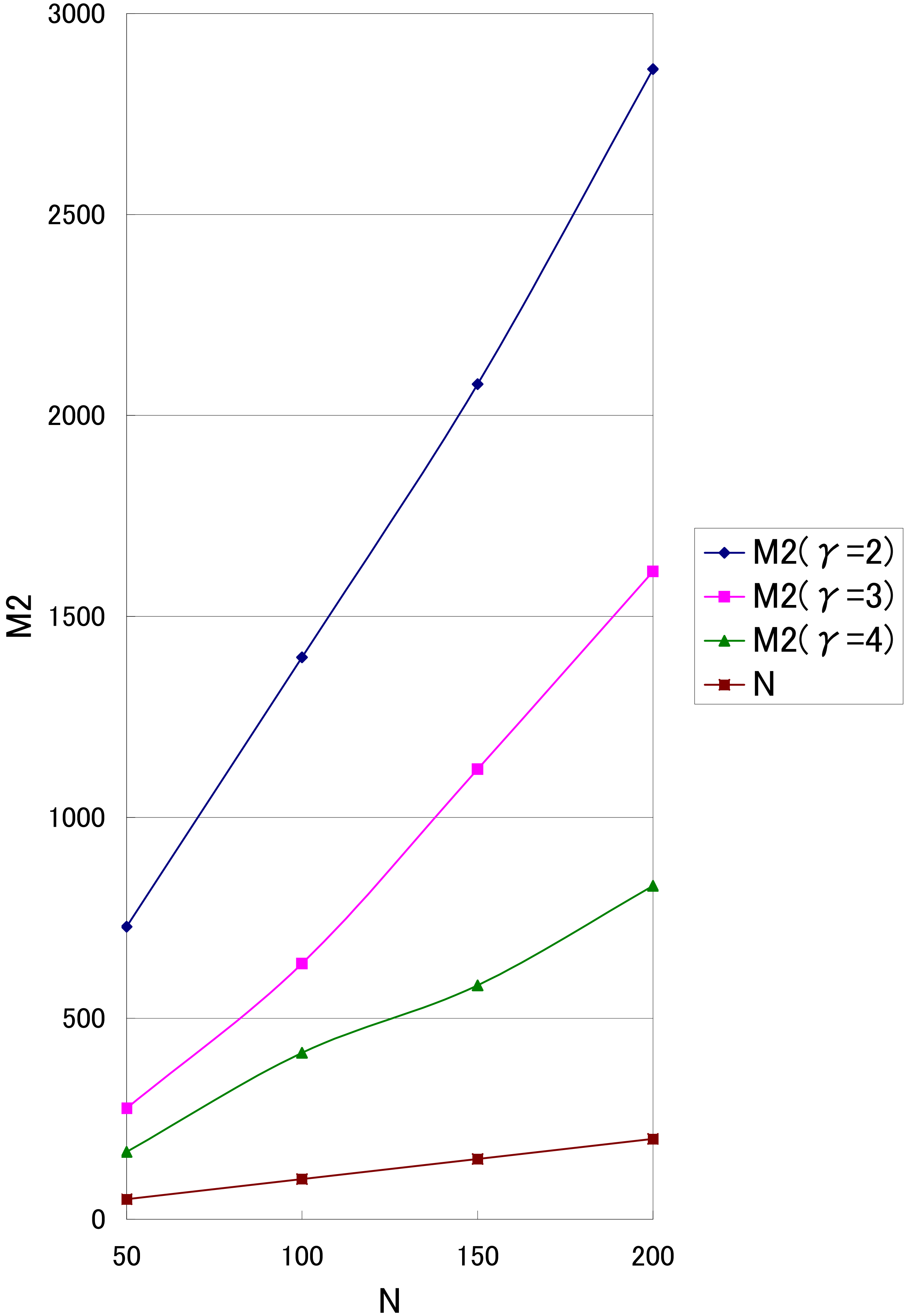
$$
\begin{aligned}
R^6[1,1]_{if} =&[A^6]_{if} + [A^4]_{if}\big(4-(k_i+k_f)\big) + [AGA]_{if}(k_i+k_p) - \{AGA, A^2\}_{if} - [A^2GA^2]_{if} \\
&+ 2[A(G^2-3G)A]_{if} + 3\sum_{j,k} a_{ij}a_{jk}a_{kf}[A^2]_{jk} - \sum_j [A^3]_{jj}\big(a_{ij}[A^2]_{jp} + [A^2]_{ij}a_{jf}\big) \\
&+2\sum_j [A^2]_{ij}[A^2]_{jf}\big(a_{ij}+a_{jf}\big) + [A^2]_{if}\big(k_i^2+k_f^2-3(k_i+k_f)+4\big) \\
&-[A^3]_{if}\big([A^3]_{ii}+[A^3]_{ff}\big) + \big([A^3]_{if}\big)^2 + \sum_j a_{ij}a_{jf}\Big(\big([A^3]_{ij}+[A^3]_{fj}\big) \\
&-[A^4]_{jj} - 2\big([A^2]_{ij}+[A^2]_{fj}\big) + [AGA]_{jj} + \big(([A^2]_{ij})^2 + ([A^2]_{fj})^2\big)\Big) \\
&+\Delta_{if}\Bigg([A^2]_{if}\big((k_i-1)(k_f-1)+1-[A^2]_{if}\big) - \big([A^3]_{if}\big)^2 + \sum_j [A^2]_{ij}[A^2]_{jf}\big(a_{ij}+a_{jf}\big) \\
&+\sum_j a_{ij}a_{jf}\Big(\big(([A^2]_{ij})^2 + ([A^2]_{fj})^2\big) - \big([A^2]_{ij}+[A^2]_{fj}\big)\Big)\Bigg) \\
&+a_{if}\Bigg([A^3]_{ff}\big(2k_f+k_i-5\big) + [A^3]_{ii}\big(2k_i+k_f-5\big) + [A^2]_{if}\big(11-3k_i-3k_f\big) \\
&\quad - 2\sum_j a_{ij}a_{jf}\Big(\big([A^2]_{ij}+[A^2]_{fj}\big)\Big)\Bigg), \\
R^6[1,2]_{if}+R^6[1,3]_{if} =& -\Delta_{if}\Bigg([A^2]_{if}\big([A^4]_{ii}+[A^4]_{ff}\big) + 4[AGA]_{if} - \{A^2, G^2-3G\}_{if} - \{AGA, A^2\}_{if} \\
&-4[A^2]_{if} - \sum_j a_{ij}a_{jf}\Bigg(\Big(([A^2]_{if})^2 + ([A^2]_{if})^2\Big) + 2\big([A^3]_{ij}+[A^3]_{fj}\big) - \big([A^2]_{ij}+[A^2]_{fj}\big)\Bigg)\Bigg) \\
&+a_{if}\Bigg(-2[A^2]_{if}[A^3]_{if} + 2[A^2]_{if}(k_i+k_f-3) + 2\sum_j a_{ij}a_{jf}\big([A^2]_{ij}+[A^2]_{fj}\big)\Bigg), \\
R^6[1,4]_{if} =&[A^3]_{if}\Delta_{if}\Big(([A^3]_{if})^2 - 3[A^2]_{if} + 2\Big).
\end{aligned}
\tag{27}
$$

$R^6[2]_{if}$, $R^6[3]_{if}$ and $R^6[4]_{if}$ are respectively given by the following expressions;

$$
\begin{aligned}
R^6[2]_{if}+R^6[3]_{if} =& a_{if}\Bigg(2[A^4]_{if} - \big(([A^5]_{ii}+[A^5]_{ff}\big) - 7\big([A^3]_{ii}+[A^3]_{ff}\big) + 22[A^2]_{ij} \\
&+ 4[A^3]_{if}[A^2]_{if} + 2\big([A^3]_{ii}k_i + [A^3]_{ff}k_f\big) + \sum_j [A^3]_{jj}\big(a_{jf}+a_{ij}\big) \\
&- 4\sum_j a_{ij}a_{jf}\big([A^2]_{ij}+[A^2]_{fj}\big) - 6\{A^2, G\}_{if} - 2[AGA]_{if} + \{A, AGA\}_{ii} + \{A, AGA\}_{ff}\Bigg), \\
R^6[4]_{if} =&a_{if}\Bigg([A^4]_{if} - [AGA]_{if} - \{A^2, G\} + 5[A^2]_{if} - \Big([A^3]_{ii}+[A^3]_{ff}\Big)\Bigg).
\end{aligned}
\tag{28}
$$

By unifying all the terms, we obtain the full expression of $R^6$. It is too long and complex that we do not describe it here. Lastly we give the expressions of Tr $R^n$ appearing in Eq. (7).

$$
\begin{aligned}
Tr(R^2) &= 0, \\
Tr(R^3) &= Tr(A^3), \\
Tr(R^4) &= Tr(A^4) - 3Tr(GA^2), +2Tr(A^2) + Tr(G^2 - G), \\
Tr(R^5) &= Tr(A^5) - 3Tr(GA^3) + 6Tr(A^3) - diag(A^3)Tr(A^2) + Ndiag(2A^3G - A^3), \\
Tr(R^6) &= Tr(A^6) + 6Tr(A^4) - 5Tr(GA^4) - 4Tr(A^3) + Tr(A^2G^2) - 6Tr(A^2G) + 4Tr(A^2) \\
&\quad + 2Tr(AGAG) - \sum_i (a_{ii})^2 - \sum_{i,j} [A^3]_{jj} a_{ij} [A^2]_{ij} + 6 \sum_{i,j} a_{ij} [A^2]_{ij} + \sum_{i,j,k} a_{ij} a_{jk} a_{ki} [A^2]_{jk}.
\end{aligned}
\tag{29}
$$



3000
2500
2000
1500
1000
500
0
50
100
150
200
M2
N
M2( γ =2)
M2( γ =3)
M2( γ =4)
N